

Scanning Guidelines for Spalding Sports and Athletic Guides
(A revised version of the Scanning Instructions dated November 3, 2000)
Preservation Reformatting Division, Library of Congress
January 22, 2001

This document shall serve as the basic guidelines for scanning the twenty-volume sub-set of *Official Indoor Baseball Guide*. Please feel free to contact Shirley Liang, when questions arise, at 202-707-5960 via phone and at slia@loc.gov via email.

I. Resolution, Format and Compression:

1. **Resolution:** Grayscale and color images shall be required at a resolution of 600 dots per inch (dpi). Grayscale images shall have a pixel depth of 8 bits-per-pixel while color images shall have a pixel depth of 24 bits-per-pixel or higher
2. **Format:** An image file shall be saved as an Intel-version (IBM byte order) TIFF file with a file extension: .tif. Files saved as such must work in IBM-compatible environment.
3. **Compression:** Uncompressed "Intel" TIFF.

II. File/Directory Name Structure:

1. File and directory names serve as pathnames in the Library's UNIX file system and will be copied en masse onto the UNIX server and used later for Web display. Since file/directory names constitute the essential part of structural metadata and their accuracy affects the search, discovery and retrieval of the scanned images and their converted texts, the file name, its continuous and sequential order as well as the hierarchical structure of file directories shall be named exactly as specified in II 2.
2. File-naming scheme:

Aggregate/itemid/scannedimagesequencenumber.tif

Example:

spalding/00238/00001.tif
spalding/00238/00002.tif
spalding/00238/00003.tif

Note:

"spalding," top level of the file directory/root directory path, is the aggregate name and provided by PRD;
"00238," second level of the file directory path, is the unique item id for each volume and provided by PRD;
"00003," third level of the file directory path, is the scanned image sequence number and shall be generated at the time of scanning.

Both the second and third level of the directory path shall contain five digits. And it is crucial to keep the scanned **page-image sequence number**, third level of the file naming structure, **continuous and complete**.

All the file names shall be in **lowercase**, because these files will be loaded onto a UNIX server.

III. Missing Pages:

1. If a leaf (two pages) is found missing, two file/place holders shall be saved for the recto and verso of the leaf to be replaced by the replacement images in the future.
2. The desired outcome is to have ready a "false page" (or target), printed with the words "Page Missing," and to scan that target-page to take the place of the missing page, and to have the correct filename for the missing page assigned to the image of the target. **Example:** If the book 00001 is missing its front cover, save two placeholders for the recto and verso of the front cover. Name the two "false-page" image files as

the following (see Scanning Guidelines II. 2): “spalding/00001/00001.tif” and “spalding/00001/00002.tif;” so replacement images can be placed into those spaces in the future.

3. (This entry applies to the twenty volumes of the *Official Indoor Baseball Guide*.) Pay special attention to the pink target sheets accompanying each volume for notes on missing pages. In case missing pages are left unnoted on the pink target sheets, placeholders shall be saved for the actual missing pages encountered during scanning.

IV. Cropping:

1. Images shall be cropped to show the entire original page and beyond the page’s edges. The reason is that the Library wishes to provide researchers with a reproduction of the entire digital item.
2. The Library wishes Acme to continue using the white background as is done with the first *Spalding Official Baseball Guide* (item 145).
3. The amount shown beyond the edge of the item shall be no less than 1 percent of the dimension of the long side image.

V. Image Orientation

In the delivered digital image, the top of the original document or page shall appear at the top of the display screen. An illustration or table in a book may be printed “sideways” (landscape) to fit the page, thus aligning the top of the page with the side of the illustration or table. In these cases, the top of the image shall be the top of the page and not the top of the illustration.

VI. Skewing

It is desirable that Acme scanning staff will make the best effort to create images with the least skew possible.

VII. MD5 Checksum

It is desirable that MD5 checksum files, stored in a separate folder, shall be added to all the volumes scanned.

VIII. Scanning Log or Read-Me File

It is desirable that scanning logs or read-me files shall be provided along with the images delivered.

IX. Delivery Media

Images shall be delivered on CD-R (write-once) disks with a gold reflective layer and compatible with all ISO 9660 standards.

X. Delivery Identification

1. A *disk name* shall be assigned to each CD-R used to deliver images. The disk name shall be an eight-digit name composed of a contractor identifier, a collection identifier and a four-digit serial number starting with 0001.
Example: acsp0001, acsp002, etc. (Acme, Spalding collection, CD-R 1 and CD-R 2)
2. The disk name shall be assigned when the disk is formatted (written) and written on the disk and its container with indelible ink.
3. Each CD-R jewel box shall be labeled with the collection name (*Spalding Sports and Athletic Guides*), disk name (acsp0001), date completed (January 21, 2001), the identifier *Library of Congress*, total number of TIFF files, dpi at which files were scanned and total file size. It is similar to what was done for *Garden and Forest* bitonal images.

XI. Image Quality Assurance

1. It is desirable that Acme shall perform quality review on all materials before delivering them to the Library.
2. The quality review shall include, but not limited to, the following activities:
 - correct file extension and format applied (uncompressed Intel TIFF)
 - complete item has been captured and proper cropping has been applied
 - images open and display properly
 - images are not skewed (if the originals are dis-bound), blurred or flawed by dust or electronic flaws
 - images orientation is correct
 - page-image file name is in continuous and sequential order (no duplicate images, no images left unscanned and no files named out of the sequential order, for instance)
 - accuracy of file-naming and directory structure
 - existence and completeness of checkmd5.fil file within each directory
 - completeness and accuracy of delivery documentation
 - delivery CD properly marked

XII. LC Quality Review and Acceptance

1. The Library will need four (4) weeks to perform inspections and to conduct tests to determine acceptance for each delivered batch. The Library will notify Acme, if more time is needed under special circumstances.
2. All unacceptable individual images shall be corrected at no additional cost to the Library.

XIII. Delivery CD-R Identification of Rework and Replacement Images

1. Rework CD Identification: Reworked images shall be delivered on *rework disks*. If a rework batch consists of a small number of images, delivery may be on a new CD-R or floppy disks or via FTP. Each rework disk shall be named and marked in a manner similar to that used for the main delivery disks (see X of this document), with the letters *rwk* added as an extension to the disk name. For example, the rework delivery CD for acsp0001 (see X. 1. of this document) shall be acsp0001.rwk. Rework disks shall not contain any previously accepted image files.
2. Replacement CD Identification: Replacement images shall be delivered on *replacement disks*. Each replacement disk shall be named and marked in a manner similar to that used for the main delivery disks, with the letters *rep* added as an extension to the name. The replacement delivery disk name shall look as follows: acsp0001.rep.

XIV. Handling

1. **Handle all materials with great care and use special caution when turning brittle pages.**
2. Move books in and out of the boxes or pamphlet binders with great care.
3. Stack unstapled pages back into their original order and line up page edges carefully so that page edges will not stick out and risk damage by the boxes they are housed in. Please be sure that pages lie flat when replacing a book back into its housing.